

The chromosomal gene structure and two mRNAs for human granulocyte colony-stimulating factor

Shigekazu Nagata, Masayuki Tsuchiya, Shigetaka Asano, Osami Yamamoto¹, Yuichi Hirata¹, Naoki Kubota¹, Masayoshi Oheda¹, Hitoshi Nomura¹ and Tatsumi Yamazaki¹

Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, and ¹New Drug Research Laboratories, Chugai Pharmaceutical Co., 3-41-8 Takada, Toshima-ku, Tokyo 171, Japan

Communicated by C. Weissmann

Two different cDNAs for human granulocyte colony-stimulating factor (G-CSF) were isolated from a cDNA library constructed with mRNA prepared from human squamous carcinoma cells, which produce G-CSF constitutively. The nucleotide sequence analysis of both cDNAs indicated that two polypeptides coded by these cDNAs are different at one position where three amino acids are deleted/inserted. When the two cDNAs were introduced into monkey COS cells under the SV40 early promoter, both of them produced proteins having authentic G-CSF activity and some difference in the specific activity was suggested. A human gene library was then screened with the G-CSF cDNA and the DNA fragment containing the G-CSF chromosomal gene was characterized by the nucleotide sequence analysis. The human G-CSF gene is interrupted by four introns and a comparison of the structures of the two G-CSF cDNAs with that of the chromosomal gene indicated that the two mRNAs are generated by alternative use of two 5' splice donor sequences in the second intron of the G-CSF gene. When the G-CSF chromosomal gene was expressed in monkey COS cells by using the SV40 enhancer two mRNAs were detected by S1 mapping analysis. Key words: alternative splicing/cDNA/chromosomal gene/human granulocyte-colony stimulating factor/S1 mapping

Introduction

The proliferation and differentiation of the progenitor cells for granulocytes and macrophages are regulated by a family of proteins called colony-stimulating factors (CSF) (Burgess and Metcalf, 1980; Metcalf, 1985). In the murine system, four CSFs have been well characterized; granulocyte-macrophage CSF (GM-CSF), granulocyte CSF (G-CSF), macrophage CSF (M-CSF) and interleukin 3 (IL-3) (Metcalf, 1985). G-CSF and M-CSF are the specific proliferating factors for granulocytes or macrophages, respectively. GM-CSF works in both the granulocyte and macrophage lineages while IL-3 stimulates the formation of a wide range of colonies consisting of not only granulocytes and macrophages but also eosinophils, megakaryocytes, erythroid and mast cells. All of these factors have been purified to homogeneity (Stanley and Heard, 1977; Ihle *et al.*, 1983; Nicola *et al.*, 1983; Sparrow *et al.*, 1985) and the gene structures for IL-3 and GM-CSF have recently been determined (Fung *et al.*, 1984; Yokota *et al.*, 1984; Gough *et al.*, 1984). Some human CSFs were also partially or homogeneously purified from the medium conditioned by human placenta, mitogen-induced T-cells or a tumor cell line

producing CSFs (Nicola *et al.*, 1979, 1985; Gasson *et al.*, 1984) and the gene structure for GM-CSF and M-CSF have been reported (Wong *et al.*, 1985; Lee *et al.*, 1985; Kawasaki *et al.*, 1985).

Recently, Nomura *et al.* (1986) established a human squamous carcinoma cell line (CHU-2) from a human oral cavity tumor which produces a high quantity of G-CSF constitutively, and the G-CSF produced by CHU-2 cells was purified to homogeneity. In a previous paper (Nagata *et al.*, 1986), we have determined the partial amino acid sequence of the purified human G-CSF, and isolated the cDNA encoding for human G-CSF by using oligonucleotide as probe. The nucleotide sequence of the cDNA was determined and the expression of the cDNA in monkey COS cells under the control of SV40 early promoter gave rise to a protein which has the authentic G-CSF activity. However during the amino acid sequence determination of the purified G-CSF protein, it was noticed that there is a definitive difference at one position between the amino acid sequence of the purified G-CSF and that deduced from the nucleotide sequence of the G-CSF cDNA. We now describe the isolation of the second G-CSF cDNA which also codes for a functionally active G-CSF protein. The protein coded by the second cDNA has a deletion of three amino acids at residues number 36–38 from the NH₂ terminus of the protein, when compared with that coded by the G-CSF cDNA isolated previously (Nagata *et al.*, 1986). The human chromosomal gene for G-CSF was then isolated from a human gene library and the gene structure was determined. The human G-CSF gene consists of five exons and a comparison of the sequences of the two cDNAs with that of the chromosomal gene has indicated that an alternative use of 5' splice donor sequence in intron 2 is responsible for the production of two different mRNAs for human G-CSF.

Results

cDNA for G-CSFb mRNA

In a previous report (Nagata *et al.*, 1986), we have described the isolation of a human G-CSF cDNA clone. The amino acid sequence deduced from the nucleotide sequence of the cDNA was identical to the sequence of the 21 NH₂-terminal amino acids of native G-CSF (Nomura *et al.*, 1986) and those of several peptides generated by chemical cleavage or enzymatic digestion of the intact G-CSF. However, when the NH₂-terminal amino acid sequence of the intact G-CSF was reinvestigated with the peptide generated with cyanogen bromide, it revealed a single definitive difference between the amino acid sequence of native G-CSF and that deduced from the cDNA. As shown in Figure 1, around amino acid position 35, three amino acids (Val-Ser-Glu) were missing on the amino acid sequence of the native G-CSF, compared with that deduced from the nucleotide sequence of G-CSF cDNA of pBRG-4 or pBRG-5 (Nagata *et al.*, 1986). This result suggested that there should exist at least two different mRNAs for G-CSF, one (G-CSFa mRNA) corresponds to the cDNA (pBRG-4 or pBRG-5) isolated previously, and the other

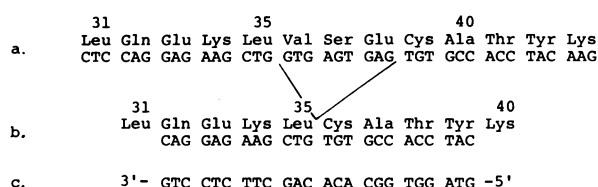


Fig. 1. Comparison of the amino acid sequence of pBRG-4 cDNA (Nagata *et al.*, 1986) and that of G-CSF purified from the conditioned medium of CHU-2 cells. (a) Nucleotide sequence and deduced amino acid sequence of pBRG-4 cDNA. The sequence of Leu-31 to Lys-43 of pBRG-4 cDNA is shown. (b) Amino acid sequence of human G-CSF. The amino acid sequence from the 31st to 40th residues from the NH₂ terminus of the purified G-CSF protein is shown together with the corresponding nucleotide sequence of pBRG-4 cDNA. (c) 24-mer oligonucleotide probe used to identify pBRV-2 cDNA.

(G-CSFb mRNA) codes for the purified protein having the amino acid sequence described in Figure 1b.

Two mechanisms are conceivable to generate more than one mRNA for human G-CSF. Like human interferon- α (Nagata *et al.*, 1980) or some other proteins (Welton-Jones and Kafatos, 1980; Kindle and Firtel, 1978), it is possible that G-CSF is coded by more than one gene. The other possibility is alternative splicing. In several cases (DeNoto *et al.*, 1981; Schwarzbauer *et al.*, 1983; Rosenfeld *et al.*, 1984; Nabeshima *et al.*, 1984; Nawa *et al.*, 1984), the alternative use of exons can generate mRNAs encoding for different polypeptide products. Since human G-CSF appeared to be encoded by a single gene (Nagata *et al.*, 1986), the latter possibility, the alternative splicing was thought to be more likely. Therefore, to isolate the cDNA coding for the G-CSF protein having the amino acid sequence shown

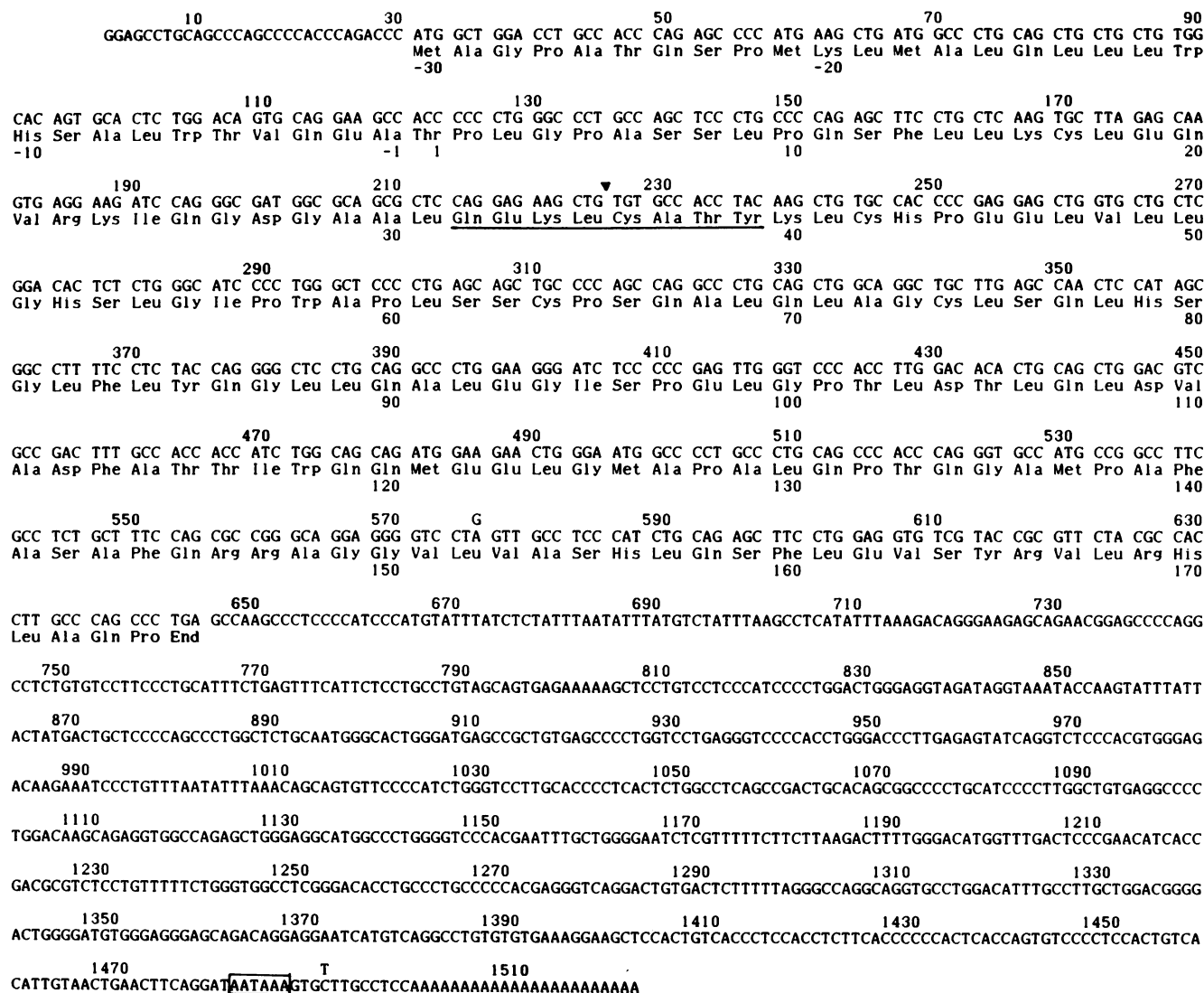


Fig. 2. Nucleotide sequence and deduced amino acid sequence of pBRV-2. Amino acids are numbered starting at Thr-1 of the mature G-CSF protein. The amino acid sequence used to make the oligonucleotide probe is underlined, and the poly(A) addition signal AATAAA (Proudfoot and Brownlee, 1976) is boxed. The sequence of pBRV-2 cDNA differs at three positions from that of pBRG-4 cDNA. At nucleotide position 225 (marked by ▼), there is a deletion of nine nucleotides and at positions 576 and 1490, pBRV-2 cDNA has A and C residues, respectively while G and T residues are found on pBRG-4 cDNA (Nagata *et al.*, 1986).

in Figure 1b, a synthetic oligonucleotide 24 bases long was prepared as a hybridization probe (Figure 1c). The nucleotide sequence of the oligonucleotide is derived from the pBRG-4 cDNA sequence (Nagata *et al.*, 1986) of the amino acid position Gln-32 to Leu-35 and Cys-39 to Tyr-42. The nine nucleotides coding for Val-Ser-Glu at the amino acid position 36–38 were removed to design the 24-mer oligonucleotide.

Approximately 300 000 cDNA clones prepared with the λ gt10 vector (Huynh *et al.*, 1985) were screened with the two different probes, one with the 32 P-labelled pBRG-4 cDNA and the other with 32 P-labelled 24-mer oligonucleotide described in Figure 1c. About 100 clones hybridized with the pBRG-4 cDNA probe, and ~20 clones were positive with both pBRG-4 cDNA and the 24-mer oligonucleotide probe. Six clones were randomly picked from those positive clones, plaque purified and designated as λ v-1 to λ v-6. The cDNAs of these clones were 1.4–1.6 kb long, as judged by *Eco*RI restriction enzyme digestion, and one of the cDNAs (λ v-2) was re-cloned at the *Eco*RI site of pBR327 (denoted as pBRV-2).

Figure 2 shows the structure of the cDNA of pBRV-2 which contains 1498 nucleotides excluding the poly(A) tract. Around amino acid position 35, exactly nine nucleotides, GTGAGTGAG, which are at the nucleotide positions 227–235 on the pBRG-4 cDNA (Nagata *et al.*, 1986), are missing in the cDNA of pBRV-2. This result confirmed that the insert of pBRV-2 is the cDNA for G-CSFb mRNA coding for the G-CSF protein, which was purified from the conditioned medium of CHU-2 cells and has the amino acid sequence described in Figure 1b. The other differences found between the nucleotide sequences of pBRG-4 and pBRV-2 are; the 5' terminus of pBRV-2 cDNA is one base shorter than pBRG-4, and the nucleotides at positions 576 and 1490 (counted on the pBRV-2 sequence) of pBRV-2 are A and C, respectively, while those of pBRG-4 are G and T. Except for these differences, the nucleotide sequence of pBRV-2 was identical to that of pBRG-4, including the 3' non-coding region which is known to diverge very quickly between homologous genes (Lomedico *et al.*, 1979). These results strongly suggest that the two G-CSF mRNAs (G-CSFa and G-CSFb) are transcribed from a single gene.

The G-CSFb mRNA can code for a protein consisting of 204 amino acids, of which 174 amino acids are for the mature protein. The total number of amino acids residues is three less than that coded by G-CSFa mRNA, and the mol. wt of the mature G-CSFb protein was calculated to be 18 671.

The structure of the human G-CSF chromosomal gene

To investigate the chromosomal gene organization and the mechanism of the generation of the two G-CSF mRNAs (G-CSFa and G-CSFb), the human chromosomal gene for G-CSF was isolated. 5×10^5 plaques of a human genomic library (a gift from Dr T. Maniatis; Lawn *et al.*, 1978), was screened with the 32 P-labelled ~600-bp *Sau*3AI fragment of pHCS-1 (Nagata *et al.*, 1986). Fifteen individual hybridizing clones were isolated from the library and DNA from seven recombinant phages was then analyzed by restriction enzyme digestion and Southern hybridization with pBRG-4 cDNA as probe. All of these DNAs had almost identical restriction maps, and the 4.2-kb *Eco*RI–*Xho*I fragment, which seemed to contain all the sequence in the cDNA of pBRG-4, was subcloned in the *Eco*RI site of pBR327 using an *Eco*RI linker.

Figure 3b shows the nucleotide sequence of the G-CSF chromosomal gene together with those of two G-CSF cDNAs.

The sequence of the G-CSF chromosomal gene and that of pBRG-4 matched completely except for a single base pair transition at nucleotide position 2066 of Figure 3b (position 1235 of Nagata *et al.*, 1986). On the other hand, the sequence of the G-CSF chromosomal gene and that of pBRV-2 differ at three positions at nucleotide positions of 1417, 2066 and 2331 of Figure 3b (positions 576, 1225 and 1490 in Figure 2b). These differences are probably due to allelic variation.

Comparison of the genomic DNA sequence with the cDNA sequences of pBRG-4 and pBRV-2 enabled us to identify the structural organization of the human G-CSF gene (Figure 3a). There are five exons in the G-CSF gene, and exons 1, 3, 4 and 5 are present in both G-CSFa and G-CSFb mRNA. On the other hand, exon 2 is used differently in G-CSFa and G-CSFb mRNA. As shown in Figure 4a, at the 5' terminus of intron 2, two donor sequences for splicing are arranged in tandem, 9 bp apart. G-CSFa mRNA utilizes the second donor sequence at nucleotide position 380/381 for splicing while G-CSFb mRNA is generated by splicing at nucleotide position 371/372. The 3' acceptor site of intron 2 at 758/759 is used in common to produce both mRNAa and mRNAb. All of the splice donor and acceptor sites conform to the GT----AG rule (Breathnach and Chambon, 1981) for nucleotides immediately flanking exon borders. Further flanking sequences are in good agreement with favored nucleotide frequencies noticed in other split genes (Breathnach and Chambon, 1981; Mount, 1982) as shown in Figure 4b. The first nucleotides of pBRG-4 (Nagata *et al.*, 1986) and pBRV-2 (Figure 2b) cDNAs are at nucleotide positions –31 and –30, respectively, and at position of –63, the sequence 5'-TATAAA-3' can be found. This sequence may correspond to the well-conserved 'Hogness box' (Breathnach and Chambon, 1981) which determines the specificity of the initiation of mRNA synthesis by RNA polymerase II (Grosschedl and Birnstiel, 1980). Since the A residues in the CA dinucleotide are generally preferred as capping sites for transcription initiation, the A at –35 can be assigned as the first nucleotide of G-CSFa and G-CSFb mRNA. From these results, we conclude that both G-CSFa and G-CSFb mRNAs are generated from a common precursor RNA as a consequence of alternative use of the splice donor sequences at the 5' terminus in intron 2.

Expression of G-CSF cDNA and the chromosomal gene in monkey COS cells

In a previous report (Nagata *et al.*, 1986), we have prepared the expression plasmid for pBRG-4 cDNA (pHGA410) and proved that the cDNA codes for a protein having G-CSF activity by expressing it in monkey COS cells. To examine whether pBRV-2 cDNA can code for a functional G-CSF, a similar expression plasmid was constructed and designated as pHGV21 (Figure 5a). The plasmids pHGA410 and pHGV21 (Figure 5a) were introduced into COS cells, and incubated at 37°C for 72 h. The media were collected and G-CSF in the media was directly, or after partial purification, assayed for CSF activity on human bone marrow cells. As shown in Table I, the supernatants of COS cells transfected either with pHGA410 or pHGV21 stimulated colony formation by using the partially purified samples, whereas no CSF activity was detected in the sample prepared from the medium of COS cells which received the vector plasmid pDKCR (Table I). Supernatants of both pHGA410 and pHGV21-derived COS cells stimulated the formation of colonies consisting almost exclusively of granulocytes (data not shown). Furthermore the G-CSF in the supernatants of COS cells transfected with either pHGA410 or pHGV21 were also active on mouse bone marrow

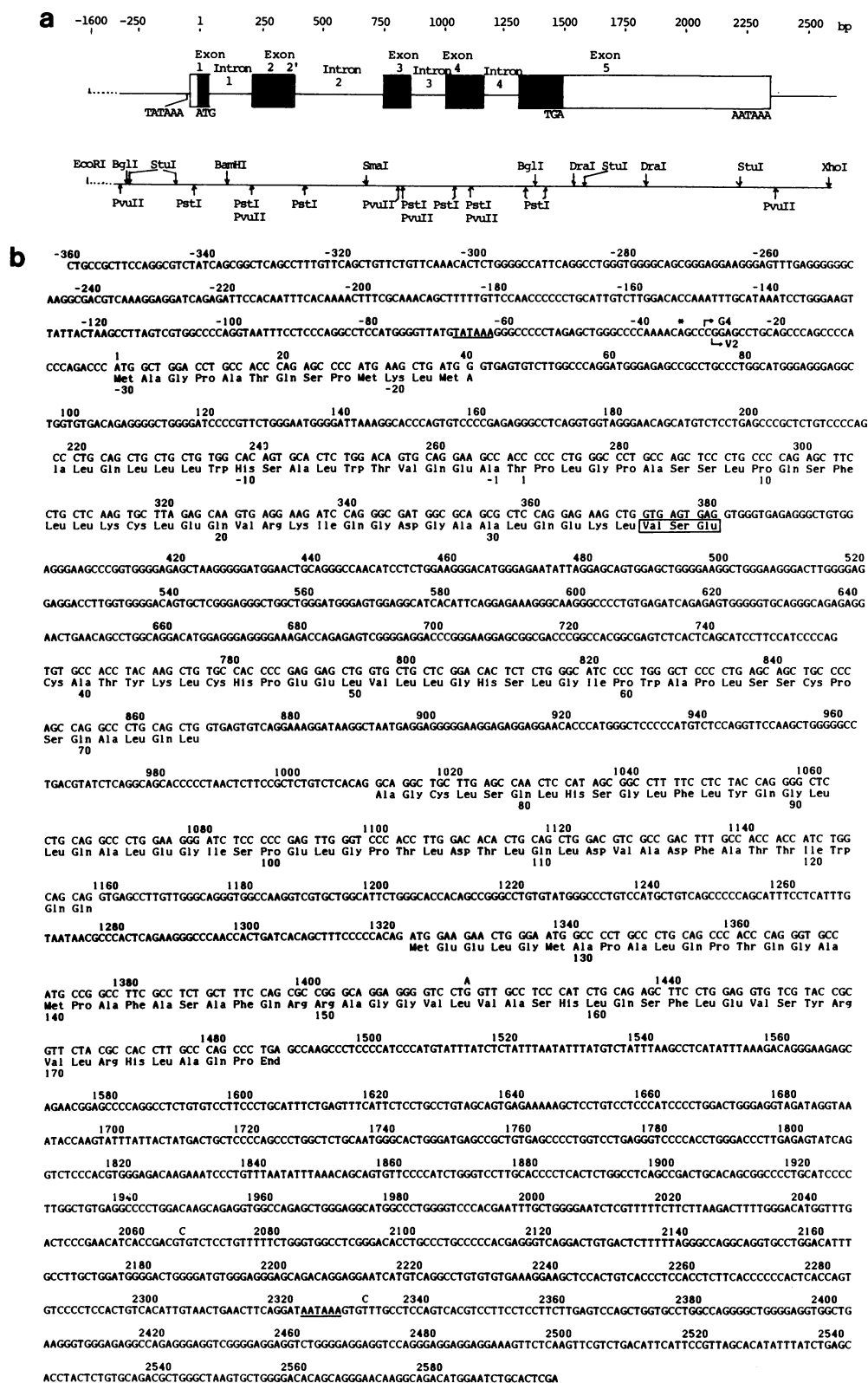


Fig. 3. Organization and nucleotide sequence of human G-CSF chromosomal gene. (a) The organization of the human G-CSF chromosomal gene. Boxes and lines between them represent five exons and four introns, respectively. The coding sequence is represented by the dark area, and the hatched region indicates intron 2' which is used as part of exon 2 on G-CSF α mRNA but not on G-CSF β mRNA. The size scale in bases is drawn above the gene; the initiation codon ATG is numbered as 1. The locations of the major recognition sites for restriction enzymes are given under the gene. (b) Nucleotide sequence of the G-CSF chromosomal gene. The coding sequence of the exons is translated and numbered from Thr-1 of the NH₂-terminal amino acid of the mature G-CSF protein. The three amino acids (Val-Ser-Glu) which are missing in the pBRV-2 cDNA are boxed. The 'TATAAA' box and 'AATAAA' polyadenylation signal are underlined. The first nucleotides of pBRG-4 and pBRV-2 are indicated, and the putative initiation site for transcription is marked by *. The nucleotide sequence of the chromosomal gene differs from that of cDNAs at three positions. At positions 1417 and 2331, pBRV-2 cDNA has A and C, respectively, while the chromosomal gene and pBRG-4 have G and T. At position 2066, the T residue of the chromosomal gene was replaced by C on the two cDNAs of pBRG-4 and pBRV-2.

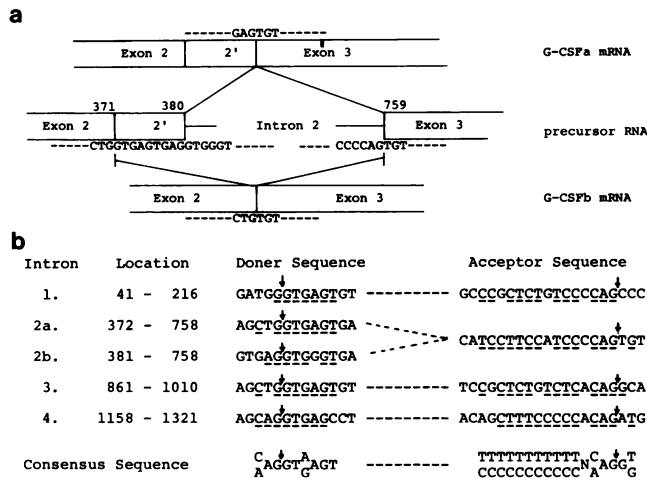


Fig. 4. Schematic representation of alternative splicing of G-CSF precursor RNA and the summary of the splice junction. (a) Alternative splicing in intron 2 of the human G-CSF gene. The structure of the splice junction in intron 2 is shown. Exons and introns are represented by open bars and lines, respectively. The numbers on the precursor RNA indicate the nucleotide position in Figure 3b. (b) Summary of introns and splice junctions of the G-CSF gene. Each row lists the introns, the nucleotide positions comprising each intron and the DNA sequences adjacent to the 5' (donor) and the 3' (acceptor) border (see Figure 3b). The junctions between exons and introns are indicated by ↓. The consensus sequences (Breathnach and Chambon, 1981; Mount, 1982) for splice site are shown on the bottom row, and the matching nucleotides on the G-CSF gene are underlined.

cells (data not shown). On the other hand, when the supernatants of COS cells were directly assayed for CSF activity, the activity was detected only in the supernatant of COS cells transfected with pHGV21 (Table I, exp. 4). Since the partial purification of G-CSF resulted in ~20 times concentration of G-CSF, these results might suggest that the two G-CSF molecules have different specific activities although the possibility of different expression of G-CSF molecules in COS cells cannot be ruled out.

As described above, in human squamous carcinoma cells of CHU-2, there are two different G-CSF mRNAs both of which code for the functional G-CSF. However, since some rearrangement of the G-CSF gene was observed in one of the alleles of CHU-2 cells (Nagata *et al.*, 1986), it can be argued that one of two mRNAs is generated from the rearranged gene. To investigate this possibility, the G-CSF chromosomal gene was expressed in monkey COS cells. 4.2 kb of *EcoRI*–*XhoI* fragment containing the human G-CSF chromosomal gene was joined at the *EcoRI* site of pMLE⁺ (Banerji *et al.*, 1981, a gift from Dr W. Schaffner) which carries the SV40 replication origin and enhancer. The resultant recombinant plasmid was designated as pMLCE3α (Figure 5b) and introduced into COS cells by the Ca²⁺ phosphate co-precipitation method. Total RNA was prepared from COS cells 48 h after transfection, and the assay for G-CSF was carried out with the medium collected 72 h after transfection. The partially purified sample from the medium contained the G-CSF activity (Table I), and the RNA was analyzed by S1 mapping (Weaver and Weissmann, 1979). As a probe, pBRG-4 was cleaved with *AhaIII*, and the 5' terminus was labelled with ³²P. About 2800 bp of the *AhaIII*–*AhaIII* fragment, of which 722 bp is the cDNA sequence and 2133 bp is pBR327 sequence, was annealed with the RNAs from COS cells or CHU-2 cells. As shown in Figure 6, two RNA-protected bands with estimated sizes of 722 and 487 bases were seen with RNA from COS cells as well as mRNA from CHU-2 cells. The sizes of these two bands correspond to those representing cDNA

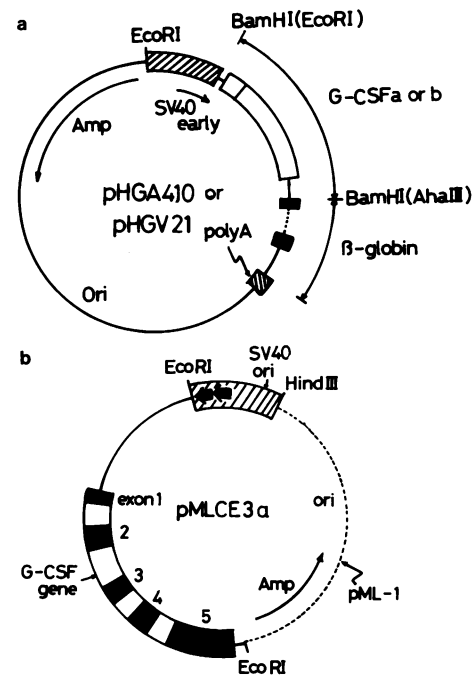


Fig. 5. Structure of the expression vector for G-CSF cDNAs and the chromosomal gene. (a) Structure of the expression vector for G-CSF cDNAs. The construction of expression plasmid pHGA410 was described previously (Nagata *et al.*, 1986), the plasmid pHGV21 was prepared in the same way as pHGA410 using 719 bp of the *EcoRI*–*AhaIII* fragment of pBRV-2 (Figure 2). For pHGV21, *Bgl*II linker was used to join the cDNA with pDKCR (O'Hare *et al.*, 1981; Fukunaga *et al.*, 1984) instead of *Bam*HI linker which was used to construct pHGA410. (b) Structure of the expression vector for the G-CSF chromosomal gene. Plasmid pMLE⁺ was constructed by ligating the 366-bp *KpnI*–*HindIII* fragment containing the SV40 replication origin with pML-1 (Lusky and Botchan, 1981) at the *EcoRI*–*HindIII* site using *EcoRI* linker. The 4.2-kb *EcoRI* fragment of pBRCE3 containing the G-CSF chromosomal gene (see Figure 2) was ligated with pMLE at the *EcoRI* site. The recombinant plasmid containing the G-CSF gene in the orientation shown in the figure was designated as pMLCE3α. The hatched area is derived from the SV40 sequence and two → in it represent the 72-bp repeats having enhancer activity (Banerji *et al.*, 1981). Dark boxes and the open boxes between them represent exons and introns of the G-CSF gene, respectively.

Table I. CSF activity in the supernatant of COS cells

	Transfected with				Control	
	pHGA410	pHGV21	pMLCE3α	pDKCR	G-CSF	Saline
Exp. 1 (a)	128,116	—	—	0	0	125,130
(b)	54, 58	—	—	0	0	—
Exp. 2 (a)	—	136,166	—	0	0	140,125
(b)	—	150,166	—	—	—	—
(c)	—	128,114	—	—	—	—
Exp. 3 (a)	—	—	18, 18	0, 0	—	18, 14
(b)	—	—	4, 2	—	—	—
Exp. 4	0, 0	130, 92	—	0, 0	—	136,116

In Exp. 1, 100 (a) and 50 μl (b), in Exp. 2, 100 (a), 30 (b) and 10 μl (c), and in Exp. 3, 100 (a) and 50 μl (b) of the partially purified sample was used for assay. In Exp. 4, 100 μl of the medium of COS cells was directly used for assay without purification. As controls, the purified G-CSF (20 ng) or saline was added to the assay mixture. Numbers of colonies consisting of >50 cells were counted on day 7.

fragments protected by G-CSFa and G-CSFb mRNA, respectively. The ratio of the 722-base fragment to the 487-base fragment was ~2:8 and 8:2, with RNA from CHU-2 cells and COS cells, respectively. This result indicates that the splice donor se-

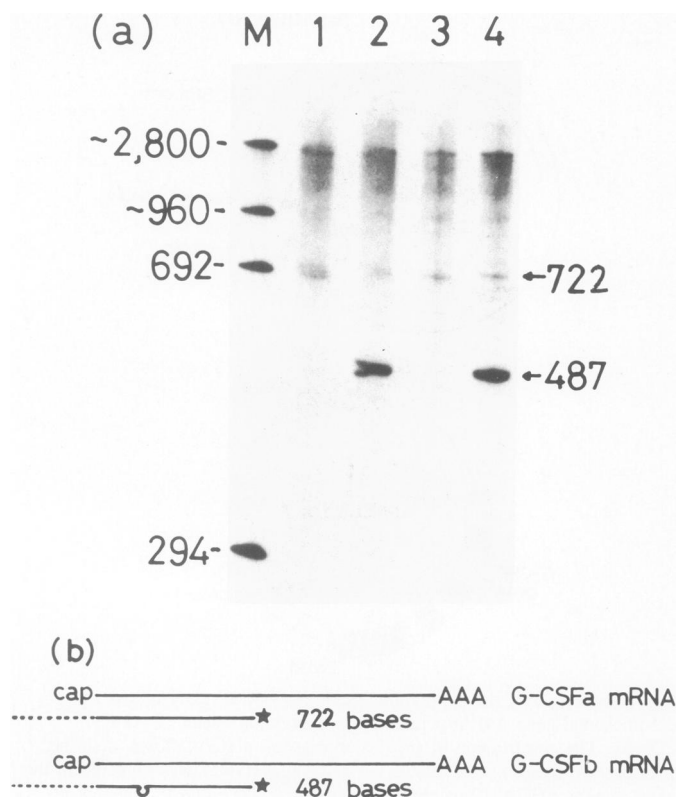


Fig. 6. S1 mapping of G-CSF mRNA. (a) Autoradiogram. lane M; size markers, 32 P-labelled *Aha*III fragment of pBRG-4 (Nagata *et al.*, 1986). The sources of RNA used for S1 mapping were; lanes 1 and 3, COS cells transfected with plasmid pMLCE3 α ; lanes 2 and 4, CHU-2 cells. Numbers on the left are the size of marker DNA in bases, and numbers on the right are the sizes of the protected cDNA. The concentrations of S1 nuclease used are: lanes 1 and 2, 200 U/ml; lanes 3 and 4, 400 U/ml. (b) Predicted duplex formation between the cDNA probe and the two mRNAs. * indicates the labelled termini and numbers are the predicted sizes of 32 P-labelled cDNA which should be protected from S1 nuclease digestion.

quences at the nucleotide positions of 371/372 and 380/381 in intron 2 are utilized in COS cells as well as in CHU-2 cells, although COS cells and CHU-2 cells have the opposite preference for the two splice donor sequences.

Discussion

The present study has revealed that the alternative use of the 5' splice donor sites in intron 2 of the human G-CSF gene generates two mRNAs which code for different polypeptide products. The alternative splicing for the generation of different mRNAs has been reported for some viral genes including SV40 (Ziff, 1980), adenovirus (Nevins, 1980) and bovine papilloma virus (Yang *et al.*, 1985) and for several cellular genes (DeNoto *et al.*, 1981; Schwarzbauer *et al.*, 1983; Rosenfeld *et al.*, 1984; Nabeshima *et al.*, 1984; Nawa *et al.*, 1984). Apart from the viral genes, most of the alternative splicings of cellular genes are mediated by the use of different exons (Rosenfeld *et al.*, 1984; Nabeshima *et al.*, 1984; Nawa *et al.*, 1984), and two examples of alternative splicing by use of different exon–intron junctions in the same exon are known (DeNoto *et al.*, 1981; Schwarzbauer *et al.*, 1983). In those cases, two or three alternative 3' splice acceptor sites were suggested to be responsible for generating the mRNAs coding for two different human growth hormones (DeNoto *et al.*, 1981) or three different rat fibronectins (Schwarzbauer *et al.*, 1983).

The G-CSF protein purified from the conditioned medium of human squamous carcinoma cell line, CHU-2, was G-CSFb having a deletion of three amino acids in the molecule, compared with the G-CSFa coded by pBRG-4 cDNA (Figure 1). This result is consistent with the S1 mapping analysis of mRNA from CHU-2 cells which showed that >80% of G-CSF mRNA in CHU-2 cells is G-CSFb mRNA (Figure 6). On the other hand, when the cloned chromosomal gene for human G-CSF was introduced into monkey COS cells, the COS cells produced mainly G-CSFa mRNA (Figure 6). It will be interesting to study whether this different splicing of G-CSF precursor RNA in CHU-2 cells and COS cells is due to some tissue-specific alternative splicing or to the rearrangement of the G-CSF gene in CHU-2 cells. The availability of the cloned G-CSF cDNA and the chromosomal gene, makes it possible to study the specific producer cells, the mechanism of the induction of the G-CSF gene and the possibility of tissue-specific alternative splicing. Furthermore, it will be necessary to determine whether two different G-CSF molecules actually exist in the human body.

The finding of two mRNAs coding for the different G-CSF polypeptide raises questions concerning the functional difference between the two G-CSF molecules and, in fact, different specific activity between the two polypeptides was suggested in the colony formation assay (Table I). In addition to the colony forming activity, G-CSF was reported to suppress the growth of some leukaemic cells such as mouse WEHI 3B-D⁺ (Nicola *et al.*, 1985), and stimulate the growth of some leukaemic cells. Therefore, to investigate the possible functional difference to two G-CSF molecules, it will be necessary to produce each G-CSF molecules on a large scale by recombinant DNA technology, and study the function of each *in vitro* and *in vivo*.

The mechanism of the alternative splicing of the G-CSF gene may provide a good model system for studying splicing of mRNA in general. The two 5' splice donor sequences in intron 2 of the G-CSF gene are arranged in tandem, and both can be utilized for splicing. This result argues against the simple scanning model (Sharp, 1981) from either the 5' or 3' side, in agreement with the results obtained by Kühne *et al.* (1983). In the current scheme for splicing, the cleavage at the 5' exon–intron junction is the first step (Ruskin *et al.*, 1984; Grabowski *et al.*, 1984). Since the two splice donor sequences in intron 2 of the G-CSF gene are separated by only nine bases and the flanking sequences of both exon–intron junctions agree well with the consensus sequence for splicing (Figure 4), it will be interesting to study what causes the preferential cleavage in one of the two splice donor sequences.

Materials and methods

Sequence determination of G-CSF protein and oligonucleotide

The G-CSF molecule (200 μ g) was carboxymethylated after reduction, and cleaved with cyanogen bromide. The peptides generated were fractionated on gel permeation h.p.l.c. on a TSK G3000 column (Toyo Soda Co.), and one of the major fragments was analyzed by the automatic protein sequencer (Applied Biosystems Inc.). Forty NH₂-terminal amino acid residues were obtained, 21 residues of which were identical to that of the intact G-CSF (Nomura *et al.*, 1986). In addition, the peptide was cleaved at the aspartic residue by incubating in 0.25 N acetic acid at 105°C for 16 h, and directly applied on the automatic protein sequencer. One of the amino acid sequences starts from Gly-28 and confirmed the sequence described in Figure 1b. A 24-mer oligonucleotide complementary to mRNA was prepared by an automatic DNA synthesizer (Applied Biosystems Inc.).

Cloning of G-CSFb cDNA

The dsDNA (Nagata *et al.*, 1986) synthesized using mRNA from CHU-2 cells was size-fractionated on 1.2% agarose gel (Low Gel Temperature, Bio Rad). DNA ranging from 1200 to 2500 bp was recovered, and a cDNA library was

constructed with the λ gt10 vector system as described (Hall and Brown, 1985). Nitrocellulose filters were prepared in duplicate and plaque hybridization (Benton and Davis, 1977) was carried out either with pBRG-4 cDNA (Nagata *et al.*, 1986) or the 24-mer oligonucleotide (Figure 1c) as probe. The pBRG-4 cDNA was nick-translated (Maniatis *et al.*, 1982) using [α - 32 P]dATP and the oligonucleotide was labelled with [γ - 32 P]ATP and T4 polynucleotide kinase. Hybridization with pBRG-4 cDNA was carried out as described (Wahl *et al.*, 1979) and hybridization with the oligonucleotide was done in $6 \times$ NET, 100 μ g/ml carrier DNA and $1 \times$ Denhardt's at 60°C (Takahashi *et al.*, 1985). The filters were washed four times in $6 \times$ SSC at room temperature, followed by washing in $6 \times$ SSC at 60°C for 1 min. The nucleotide sequence was determined by the dideoxynucleotide chain termination method after subcloning into M13 mp8 or mp9 (Messing, 1983).

Cloning of the chromosomal gene for human G-CSF

5×10^5 plaques of λ Charon 4A carrying a human DNA fragment were screened by a nick-translated ~ 600 bp of the *Sau*3AI fragment of pHCS-1 (Nagata *et al.*, 1986) as described (Wahl *et al.*, 1985). Fifteen clones gave a positive result and the recombinant DNAs of seven clones were characterized by restriction mapping and Southern hybridization with 32 P-labelled pBRG-4 cDNA. In all seven DNAs, ~ 8 kb of *Eco*RI fragment was hybridized with 32 P-labelled pBRG-4 cDNA and this fragment was subcloned at the *Eco*RI site of pBR327 (denoted as pBRCE). The fine restriction map was constructed with pBRCE and the *Eco*RI-*Xho*I DNA fragment (4.2 kb) was further subcloned at the *Eco*RI site of pBR327 by using *Eco*RI linker. The resulting plasmid was designated as pBRCE3 and nearly 3 kb of the nucleotide sequence of the G-CSF gene region was determined by the dideoxynucleotide chain termination procedure using phage M13 derivatives (Messing, 1983).

S1 mapping

8×10^6 COS cells (Gluzman, 1981) (on four 9 cm-plates) were transfected with 80 μ g of plasmid pMLCE3 α (Figure 5b), and total RNA was prepared (Chirgwin *et al.*, 1979) 48 h after transfection. As a probe, pBRG-4 was cleaved with *Aha*III, labeled at its 5' end with [γ - 32 P]ATP and T4 polynucleotide kinase. A 2.8-kb *Aha*III fragment containing 722 bp of the 5' portion of G-CSF cDNA and 2133 bp of pBR327 was isolated by an agarose gel electrophoresis and used as a probe. The probe (1.5×10^5 c.p.m., 2.8×10^6 c.p.m./ μ g DNA) was denatured and mixed with 2 μ g of mRNA prepared from CHU-2 cells or 20 μ g of total RNA from COS cells. Hybridization was carried out at 45°C for 15 h, and digested with S1 nuclease (P.L. Biochemicals) according to the protocol of Weaver and Weissmann (1979). The products were electrophoresed through an 8.3 M urea/4% polyacrylamide gel and autoradiographed.

Transfection of COS cells and in vitro colony formation assay

2×10^6 COS cells in 10 ml of medium were transfected with 20 μ g of plasmid DNA by the DEAE-dextran method (Lee *et al.*, 1985) or the Ca^{2+} phosphate co-precipitation method (Fukunaga *et al.*, 1984) (exp. 3). 40 ml of medium was collected 72 h after transfection, the G-CSF in the medium was partially purified and dissolved in 2 ml of Iscove's medium as described (Nagata *et al.*, 1986). The assay for G-CSF activity was carried out using human non-adherent bone marrow cells (Nagata *et al.*, 1986).

Acknowledgements

We thank Drs T.Maniatis and W.Schaffner for a human gene library and pMLE $^{+}$, respectively. We also thank Drs Y.Kaziro, S.Miwa, H.Ogawa, S.Hata, M.Ono, H.Okazaki and T.Hayashi for helpful discussions and encouragement.

References

- Banerji, J., Rusconi, S. and Schaffner, W. (1981) *Cell*, **27**, 299–308.
- Benton, W.D. and Davis, R.W. (1977) *Science*, **196**, 180–182.
- Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
- Burgess, A.W. and Metcalf, D. (1980) *Blood*, **56**, 947–958.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry*, **18**, 5294–5299.
- DeNoto, F.M., Moore, D.D. and Goodman, H.M. (1981) *Nucleic Acids Res.*, **9**, 3719–3930.
- Fukunaga, R., Sokawa, Y. and Nagata, S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5086–5090.
- Fung, M.C., Hapel, A.J., Ymer, S., Cohen, D.R., Johnson, R.M., Campbell, H.D. and Young, I.G. (1984) *Nature*, **307**, 233–237.
- Gasson, J.C., Weisbart, R.H., Kaufman, S.E., Clark, S.C., Hewick, R.M., Wong, G.G. and Golde, D.W. (1984) *Science*, **226**, 1339–1342.
- Gluzman, Y. (1981) *Cell*, **23**, 175–182.
- Gough, N.M., Gough, J., Metcalf, D., Kelso, A., Grail, D., Nicola, N.A., Burgess, A.W. and Dunn, A.R. (1984) *Nature*, **309**, 763–767.
- Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) *Cell*, **37**, 415–427.
- Grosschedl, R. and Birnstiel, M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1432–1436.
- Hall, A. and Brown, R. (1985) *Nucleic Acids Res.*, **13**, 5255–5268.
- Huynh, T., Young, R.A. and Davis, R.W. (1985) In Glover, D.M. (ed.), *DNA Cloning Techniques — A Practical Approach*. IRL Press, Oxford and Washington, D.C., pp. 49–78.
- Ihle, J.N., Keller, J., Oroszlan, S., Henderson, L.E., Copeland, T.D., Fitch, F., Prystowsky, M.B., Goldwasser, E., Schrader, J.W., Palaszynski, E., Dy, M. and Lebel, B. (1983) *J. Immunol.*, **131**, 282–287.
- Kawasaki, E.S., Ladner, M.B., Wang, A.M., Arsdell, J.V., Warren, M.K., Coyne, M.Y., Scheickart, V.L., Lee, M.-T., Wilson, K.J., Boosman, A., Stanley, E.R., Ralph, P. and Mark, D.F. (1985) *Science*, **230**, 291–296.
- Kindle, K.L. and Firtel, R.A. (1978) *Cell*, **15**, 765–778.
- Kühne, T., Wieringa, B., Reiser, J. and Weissmann, C. (1983) *EMBO J.*, **2**, 727–733.
- Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell*, **15**, 1157–1174.
- Lee, F., Yokota, T., Otake, T., Gemmel, L., Larson, N., Luh, J., Arai, K. and Rennick, D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4360–4364.
- Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) *Cell*, **18**, 545–558.
- Lusky, M. and Botchan, M. (1981) *Nature*, **293**, 79–81.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY.
- Messing, J. (1983) *Methods Enzymol.*, **101**, 20–78.
- Metcalf, D. (1985) *Science*, **229**, 16–22.
- Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
- Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Ogata, K. (1984) *Nature*, **308**, 333–338.
- Nagata, S., Mantei, N. and Weissmann, C. (1980) *Nature*, **287**, 401–408.
- Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O., Hirata, Y., Kubota, N., Oheda, M., Nomura, H. and Ono, M. (1986) *Nature*, **319**, 415–418.
- Nawa, H., Kotani, H. and Nakanishi, S. (1984) *Nature*, **312**, 729–734.
- Nevins, J.R. (1980) *Cell*, **28**, 1–2.
- Nicola, N.A., Metcalf, D., Johnson, G.R. and Burgess, A.W. (1979) *Blood*, **54**, 614–627.
- Nicola, N.A., Metcalf, D., Matsumoto, M. and Johnson, G.R. (1983) *J. Biol. Chem.*, **258**, 9017–9023.
- Nicola, N.A., Begley, C.G. and Metcalf, D. (1985) *Nature*, **314**, 625–628.
- Nomura, H., Imazeki, I., Oheda, M., Kubota, N., Tamura, M., Ono, M., Ueyama, Y. and Asano, S. (1986) *EMBO J.*, in press.
- O'Hare, K., Benoist, C. and Breathnach, R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1527–1531.
- Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature*, **263**, 211–214.
- Rosenfeld, M.G., Amara, S.G. and Evans, R.M. (1984) *Science*, **225**, 1315–1320.
- Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell*, **38**, 317–331.
- Schwarzbauer, J.E., Tamkun, J.W., Lemischka, I.R. and Hynes, R.O. (1983) *Cell*, **35**, 421–431.
- Sharp, P.A. (1981) *Cell*, **23**, 643–646.
- Sparrow, L.G., Metcalf, D., Hunkapiller, M.W., Hood, L.E. and Burgess, A.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 292–296.
- Stanley, E.R. and Heard, P.M. (1977) *J. Biol. Chem.*, **252**, 4305–4312.
- Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. and Matsubara, K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1931–1935.
- Wahl, G.M., Stern, M. and Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683–3687.
- Weaver, R.F. and Weissmann, C. (1979) *Nucleic Acids Res.*, **7**, 1175–1193.
- Welton-Jones, C. and Kafatos, F.C. (1980) *Nature*, **284**, 635–638.
- Wong, G.G., Witek, J.S., Temple, P.A., Wilkens, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A. and Clark, S.C. (1985) *Science*, **228**, 810–815.
- Yang, Y.-C., Okayama, H. and Howley, P.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1030–1034.
- Yokota, T., Lee, F., Rennick, D., Hall, C., Arai, N., Mosman, T., Nabel, G., Cantor, H. and Arai, K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1070–1074.
- Ziff, E.B. (1980) *Nature*, **287**, 491–499.

Received on 20 December 1985